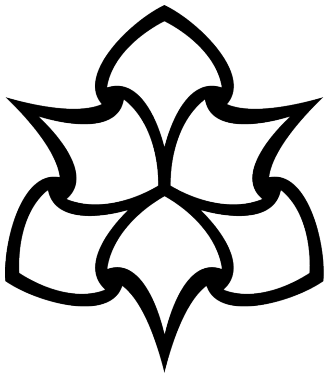

Manchester Metropolitan University

Archer, Dawn ORCID logoORCID: https://orcid.org/0000-0002-4547-6518 and Findlay, Alison (2020) Keywords that characterise Shakespeare's (anti)heroes and villains. In: Voices of the Past and Present - Studies of Involved, Speech-related and Spoken. John Benjamins Publishing Company, pp. 31-46. ISBN 9789027207654

Downloaded from: https://e-space.mmu.ac.uk/626624/

Version: Accepted Version

Publisher: John Benjamins Publishing Company

DOI: https://doi.org/10.1075/scl.97.03arc

Please cite the published version

https://e-space.mmu.ac.uk

Keywords that characterise Shakespeare's (anti)heroes and villains

Dawn Archer and Alison Findlay

Abstract

This paper undertakes a keyword analysis of seven Shakespearean characters: Titus, Tamora, Aaron, Lear, Edmund, Macbeth and Lady Macbeth. The paper discusses how, once contextualised, these keywords provide useful insights into their feelings/thoughts towards others, events, motivations to act, etc. In terms of findings, only Aaron denotes his "villainy" directly. Tamora, in contrast, draws upon a keyword that is denotatively positive; in context, though, "sweet" reveals her womanly wiles. "Weep", for Lear, and "legitimate" and "base", for Edmund, problematize their status as (one-dimensional) villains. Macbeth and Lady Macbeth draw upon grammatical keywords, "if" and "would" in ways that signal something about their (deteriorating) emotional and social positions as much as their villainous intentions.

1. Introduction

Professor Merja Kytö is well known for both her interest in "involved" texts – spoken/speech-related, historical and contemporary – and also her work in ensuring others have access to rich resources that can be interrogated using corpus linguistic techniques. In line with the above, this paper draws upon a new resource, developed as part of Lancaster University's AHRC-funded *Encyclopedia of Shakespeare's Language Project*, which allows researchers to explore Shakespeare's plays using statistical keyword methods. We will demonstrate how this technique can benefit – by confirming/refuting or advancing – existing literary understandings of Shakespearean depictions of villainy (Sections 4 – 7.2). Three of the seven Shakespearean characters under analysis – Titus, Tamora and Aaron – are taken from an early play: *Titus Andronicus* (1594). The remaining four – Lear, Edmund, Macbeth and Lady Macbeth – appear in two of the later tragedies: *King Lear* (c.1605 but revised for the Folio) and *The Tragedy of Macbeth* (c.1606).

Even quick guides like Quennell and Johnson's *Who's Who in Shakespeare* (2013: 1) indicate the complexity of these particular characters. They note of Aaron, "the forebear of other Shakespearean villains" (ibid.), that he is simultaneously:

> [..] a heartless Machiavel, an advocate of 'policy and stratagem', and 'chief architect and plotter' of the tragic events [of the play]; the evil Moor of Christian tradition…distinguished by…cruelty; [and] above all…the direct descendent of the figure of Vice in the medieval morality plays.

However, they go on to point out how his behaviour towards his son humanizes him beyond the personified figure of Vice (ibid). Quennell and Johnson (ibid: 64) describe Edmund as "a witty and attractive villain" who is nonetheless less guilty than the figure on which Shakespeare based him. Longer studies such as Charney (2012: 100) are less flattering, though, describing Edmund as being "without much scruple" and "cunning like Iago". Charney spares Lady Macbeth the *villain* label, in spite of conceding she shares her husband's "murderous and savage thoughts" (ibid: 86). Macbeth, in contrast, is argued to have "a special status in Shakespeare as a villain-hero", in part because of how he "agonizes…over his ill-doing" (ibid). We discuss the extent to which the keyword results for these characters confirm them as (anti)heroes or villains following our description of the resource used in this particular study (Section 3) and the keyword methodology adopted (Section 4). We begin, however, with an outline of similar Shakespeare-focussed keyword studies within the pragma-stylistic tradition, as a means of situating our work (Section 2).

2. Background

The use of corpus linguistic approaches to analyse Shakespeare is now well established. Studies range from fine-grained investigations of particular characters (Archer and Bousfield 2010; Culpeper 2002, 2009) to investigations across the whole body of plays, exploring themes (Archer et al. 2009 on love) or specific language features (Beatrix Busse 2006 on vocatives; Ulrich Busse 2002 on second-person pronouns; Culpeper and Oliver this volume on pragmatic noise). As well as adding some much-needed empirically based findings to the large and long-established body of qualitative literary critical work, these quantitative studies have provided useful insights into the way Shakespeare used language to construct different types of individuals, settings and plots. Culpeper (2002: 21) has suggested, for example, that the Nurse in *Romeo and Juliet* is "dispositionally emotional" (i.e., affected by and reacting to the traumatic events of the play) based on her use of the keywords "god", "warrant", "faith", "marry" and "ah". Two of Juliet's grammatical keywords – "if" and "yet" – are similar in that they are occasioned by the unfolding events. This study builds on such work in the pragma-stylistic tradition. Our aim is to analyse the words spoken by each character, in context, paying particular attention to why they are spoken and to whom. While this reduces each 'character' to a collection of words, we demonstrate how the keyword approach can be used, nonetheless, to explore the dramatized expression of their feelings and thoughts, and from this, their potential motivations (Archer and Lansley 2015). We explain our methodology in more detail in Section 4, after describing the resource drawn upon in this study.

3. Resource drawn upon

The *AHRC-funded Encyclopedia of Shakespeare's Language Project* (Grant Ref: AH/N002415/1) uses *"*c*omputers* to identify patterns of use across Shakespeare's works" (Culpeper, forthcoming) that can be difficult to detect otherwise. This paper draws on the project's core dataset*,* namely, electronic versions of the thirty-six plays of the *First Folio*

(1623) plus *Pericles* and *The Two Noble Kinsmen* from *Quarto 1* (downloadable from the Internet Shakespeare Editions [ISE] website). Project-focussed enhancements made to the electronic plays are explained in detail in Culpeper and Oliver (this volume) and so will not be discussed here beyond highlighting two, which are specifically designed to improve the accuracy of results derived from using corpus linguistic techniques like keyword analysis (Baron and Rayson, 2008). First, every original spelling within each play has been checked and, when relevant, regularised manually, according to the criteria in Culpeper (forthcoming), aided by VARD (http://ucrel.lancs.ac.uk/vard/about/). Second, each play text has been annotated using a customised instance of the CLAWS4 part-of-speech tagging software (Garside and Smith 1997), based on a variant of the C6 tagset designed to account for the language of the period. In brief, seventeenth-century vocabulary has been added to the tagger's twentieth-century lexicon, and tags added to the tagset, so we can capture the second-person singular pronouns *thou*, *thee*, etc., and thus achieve better verb agreement (with *dost*, *didst*, etc.). This is necessary because the default CLAWS tagger achieves only 89% accuracy on Shakespearean text even *with* spelling regularisation (Rayson et al., 2007). Whilst the aforementioned enhancements have raised this, it is not to CLAWS's 95-97% success rate for present-day English. In consequence, the part-of-speech annotation has also been manually post-edited to correct all errors at the major wordclass (verb/noun/adjective/etc.) level. Whilst the possibility of human errors remain, accuracy is thus "as close to 100%" as can be achieved at this time (Culpeper, forthcoming).

4. Keyword approach adopted

Previous studies like Culpeper (2002, 2009) have tended to draw on one statistical measure when determining keywords and what they might tell us about a particular character (and/or their relationships with others). We follow the approach of the Shakespeare Language project,

however, and draw on a cutting-edge three-step process. The first measure, Log-Likelihood (LL), is an indicator of *statistical significance*, that is, how much evidence we have for a given difference between two wordlists (in our case, the wordlist for a character only when compared with the wordlist for the play in which they appear, minus their turns). The results presented here are filtered using a LL cut-off of 6.63, meaning each keyword has a minimum confidence level of 99%. The second measure, *Log Ratio* (LR), is an *effect size* statistic used to sort the keyword list such that the quantitatively largest differences are at the top of that list (http://cass.lancs.ac.uk/log-ratio-an-informal-introduction/). Because LR is a binary logarithm of the ratio of relative frequencies, each increase by 1 indicates a *doubling* of how many times *more* frequent the word is, with respect to a particular character's wordlist, when compared against the *full play* wordlist (minus their turns). In order to build the analysis on only the most prominent differences among those shown to have a sufficient evidence base (using the LL filter), this paper reports on keywords with an LR ranging between 1 and 7 (making them twice to sixty-four times more frequent in the relevant character's wordlist, relatively speaking). We have also restricted the analysis to keywords with a minimum frequency of 5, on the understanding that a qualitative interpretation of a word's use in a character's speech is difficult when such words occur less than this. Taking significance *and* effect size into account, as well as a minimum frequency filter, tends to generate less keywords for consideration but also ensures both *qualitative* and *quantitative* robustness. This is important, given that keyword generation is a first stage of a two-stage approach, which also involves checking the concordance lines of each keyword as a means of determining their use in context. In our case, we are particularly interested in keywords that divulge something about the seven characters' villain vs. hero status. As such, Section 5 initially reports all generated keywords. We then select so that we can discuss specific keywords (from Table 1) in their context-of-use.

5. Keyword results for the seven characters

Table 1 shows those keywords (of LL6.63+) that occur five times or more in the seven characters' wordlists, and are ordered according to their LR values (provided in brackets, following each keyword's frequency).

Table 1: Keywords for the seven Shakespearean characters

AARON	**gold** **villainy** **black** **keep** **set** **empress**	(8/6.6) (5/4.92) (6/3.6) (7/2.82) (5/2.6) (14/1.65)	TAMORA	**ear** **fear** **Andronicus** **sweet** **revenge** **at**	(5/2.7) (5/2.51) (15/2.09) (11/1.79) (9/1.77) (13/1.46)
TITUS	**sea** **eat** **ha** **girl** **service** **drink** **sirrah** **Publius** **read** **get** **Marcus** **hold** **mine** **earth** **tribune** **tear(s)** **come** **these** **they/them**	(7/5.09) (5/4.61) (5/4.61) (5/3.61) (5/3.61) (5/3.61) (5/3.61) (8/3.28) (7/3.09) (6/2.87) (29/2.68) (8/2.28) (10/2.02) (10/2.28) (12/1.87) (25/1.68) (44/1.25) (28/1.5) (63/1.03)			
KING LEAR	**Dower** **ha** **weep** **Regan** **cause** **kill** **boy** **daughter** **her** **she**	(5/5) (10/4) (8/3.67) (13/3.67) (7/2.9) (7/2.48) (10/2.41) (29/1.89) (26/1.57) (42/1.23)	EDMUND	**Legitimate** **base** **brother** **sword** **business** **by** **father**	(5/6.55) (5/6.55) (11/2.88) (6/2.28) (5/2.38) (17/1.34) (14/1.23)
MACBETH	**Tomorrow** **born** **till** **&/and** **blood** **fear** **if**	(8/5/09) (7/3.9) (14/2.31) (10/2.09) (14/1.73) (23/1.22) (23/1.15)	LADY MACBETH	**Without** **bed** **would** **your**	(6/2.52) (5/2.45) (15/1.38) (27/1.14)

As is clear from these results, very few of the keywords (which are spoken *by* the characters rather than *about* them) see characters self-identify as villains. Only Aaron has "villainy" as a keyword, although Edmund's keyword "base" draws on the original, opprobrious meaning of "villain" as a "low-born, base-minded rustic" (*OED* 1a). Only a few of the other keywords - "blood", "kill", "fear", "revenge" – are associated with villainous behaviour. The second stage of our methodology – qualitative analysis of the context-of-use for a selection of keywords from Table 1 - is thus crucial to establish how each character's usage of a word helps define them in relation to heroism and villainy.

If we view the above collectively, we see evidence of both keywords to do with *aboutness* (i.e., content relating to the plots) and also keywords that are more grammatical in nature, but reveal potential character traits. Pronouns feature as keywords for three of the seven (Titus, Lear and Lady Macbeth) and, as the forthcoming sections reveal, allude to their (often tumultuous) relationships with others. The characters from *King Lear* have kinship terms as keywords ("daughter", "brother"), in addition, providing our first signal that family divisions act as a driver for the villainy of this particular play (see Section 7). Other "grammatical keywords" (Culpeper, 2009) of note ("if", "till", "without", "would") highlight the anxious state of the two *Macbeth* characters (see Sections 8.1 and 8.2). Macbeth also uses the emotion-related term, "fear", more than other characters from his respective play (statistically speaking). "Fear" is a keyword for Tamora, in addition, but (*contra* Macbeth) is something she deliberately causes others to feel (cf. Sections 6.2 and 8.1). The patriarchs, Titus and Lear, also draw on similar emotion-related words to the other: "tear(s)" and "weep". In each case, they allude to their heartfelt despair, occasioned by the treatment of others, but it is Lear alone who comes to

recognise their part in precipitating that treatment and demonstrating remorse in consequence (cf. Sections 6.3 and 7.1).

6. Discussion of the *Titus Andronicus* characters

Titus Andronicus foregrounds villainy through its reworking of the genre of revenge tragedy. A victorious Titus returns to Rome, having taken Tamora, Queen of the Goths, her sons, and Moorish lover prisoner. When Titus offers Tamora's eldest son as a sacrifice, Tamora schemes to marry Saturninus, new Emperor of Rome, so that, aided by Aaron, she can take revenge on Titus and his family.

6.1 Aaron

Multiple instances of two keywords – "black" and "villainy" – confirm Quennell and Johnson's (2013: 1) depiction of Aaron as "the evil Moor of Christian tradition". Aaron chooses wickedness over goodness to the point of delighting in it (like the Vice figure from medieval drama). Having tricked Titus into chopping off his hand, for example, he informs spectators: "how this *villainy* / Doth fat me with *the* very thoughts of it, / Let fools do good, and fair men call for grace, / Aaron will have his soul *black* like his face" (3.1). Although "villainy" has no statistical collocates, he uses the keyword alongside other negative terms, for example, "rape and villainy" (2.1), "villainy and vengeance" (2.1) and "Mischief, Treason, Villainies" (5.1). Aaron's identity as a "slave" (5.1. and 5.3), moreover, makes him self-identify with the opprobrious associations of the word's original meaning as low-born. His social aspirations are demonstrated by the co-occurrence of his keywords "villainy" and "gold".

Assuming kinship with the spectators, Aaron informs them that "To bury so much *Gold* under a tree" (2.3) may seem illogical, but in fact, "this *gold* must coin a stratagem, / Which cunningly

effected, will beget / A very excellent piece of *villainy*" (2.3). He affectionately addresses the "sweet *Gold*" (2.3) as an accomplice in his plot to implicate two of Titus's sons in the murder of Bassianus, revealing the "bag of *gold*" as evidence (2.3). Aaron's social ambition is further evidenced in the use of keywords "villainy" and "gold" with reference to Tamora. Aaron concedes his "*Empress*" has a "sacred wit / To *villainy* and vengeance' that is a match for his (2.1). He likens her to the "*golden* sun", but goes on to claim she is a slave to his love, allowing him to "mount her pitch" (2.1). Although Aaron's legacy (as a figure of Vice) explains his delight in plotting, the superficial rationale of "ambition and vague desire for revenge" is more complex than Quennell and Johnson suggest. His paternal affection shows that he is more nuanced than a stereotypical character denoting evil. Indeed, his use of the terms "black"' as well as "slave" (4.2 and 5.1) to address his newborn son challenges early modern cultural assumptions about blackness and villainy. He asks "is *black* so base a hue?", when his son is referred to as a devil and threatened with death (see also White 1997). He also suggests "Coal-*black* is better than another hue" when challenging Chiron and Demetrius to recognise their step-brother as their equal. Aaron's keywords, when considered collectively, thus suggest that even his refusal to repent for the "thousand dreadful things" he has done against the Romans and wish to do "a thousand more" (5.1) may be a race-specific desire for revenge against the racist hegemony that labels him a villain. Simply put, it is more than simply a reversion to a Vice-like role as "the personification of evil" (Quennell and Johnson ibid: 1).

6.2 Tamora

When viewed in their context-of-use, three of Tamora's keywords – "sweet", "ear" and "fear" – hint at her feminine wiles. After being made Empress, she appeals to her "*sweet*" Emperor to "pardon what is past" before encouraging him and Titus to come together (1.1). Spectators' suspicions about her are confirmed, however, when she confides to Saturninus she will "find a

day to massacre…all" the Andronici. Her next public display of sweet-talking, when she advises her "*sweet* emperor" they "must all be friends", assuring her "*sweet* heart" she "will not be denied" (1.1), is thus blatantly deceptive. The "revenge of the villains", as Bowers (2015: 112) calls it, is not a single act on Tamora's part. It requires, instead, her complicity in her "sweet" Moor's maiming of Titus and his kin. That Tamora is as deliberately villainous as Aaron becomes evident in turns where she brags about her sweet-talking abilities to Saturninus, telling him she "will enchant the old Andronicus, / With words more *sweet*, and yet more dangerous / Than baits to fish, or honey-stalks to sheep" (4.4). "Ear" is significant in this regard too, with Tamora claiming she "can smooth and fill his aged *ear*, / With golden promises, that were his heart / Almost Impregnable, his old *ears* deaf, / Yet … both *ear* and heart" would "obey her tongue" (4.4). Whilst the keyword "fear" might seem appropriate to a prisoner of war, their context-of-use reveal a fearless rather than fearful Tamora. Three of the five instances collocate with "not". Two "fear not" instances occur at points where Tamora is attempting to deceive others (in 1.1 and 2.3), by putting them at their ease. The third instance admonishes Saturninus at the point he fears a public uprising (4.4). Tamora also draws upon the keywords "fear" and "ear", menacingly, when disguised as Revenge, telling Titus: "There's not a hollow Cave or lurking place … / Where bloody Murder or detested Rape, / Can couch for *fear*, but I will find them out, / And in their *ears* tell them my dreadful name" (5.2).

6.3 Titus

Titus's keywords point to him being a character of extremes: a hero and anti-hero/villain. His absolute loyalty to Roman tradition and "service" (a keyword for Titus) help to explain his belief that sacrifices - like slaying Alarbus - have to be made regardless of the cost, if it ensures his own slain "brethren" can "rest" in "eternal sleep" (1.1). The keyword "mine" captures both Titus's (at times competing) attempts to control and protect the Andronicus family of which he

is head, and his inseparable link to Rome however badly it treats him. He draws on "mine" when swearing allegiance to the newly-elected Emperor, Saturninus: "I *hold* me Highly Honoured of your Grace, / And here in sight of Rome to Saturnine / …do I Consecrate, / My Sword my Chariot… / *Mine* Honour's Ensigns humble at thy feet" (1.1). Such loyalty perpetuates additional sacrifices for Titus, including rejecting his disobedient son Mutius as "*no* son of *mine"* before slaying him. Titus's keyword, "tears", occurs twenty-five times, and is a prime example of how keywords can allude to characters' emotional states. In Titus' case, we see him forgo the Roman stoicism that meant he "never wept" (3.1) to the point of being overwhelmed on learning of Lavinia's mutilation. He likens his "girl" (a keyword that Titus only uses in reference to Lavinia) to "the weeping welkin" and himself, "the *earth*": before lamenting how "*earth* with her continual *tears*" has "Become a deluge, overflow'd and drown'd" (3.1). The keywords related to weeping most clearly exemplify Titus as tragic hero rather than villain. Receiving the heads of Martius and Quintus, alongside his own severed hand, marks a change for him, however, and leads to "our fearless hero brutally exact[ing] revenge upon [his] equally vicious opponents", to quote McDonald (2000: 5). Laughing hysterically, Titus declares "I have not another *tear* to shed" (3.1) – hence the keyword, "ha". The following scenes then see him transcend from grief through madness into a single-minded revenge, as he comes to appreciate the full extent of Tamora and Aaron's plot against him. The tribal nature of this revenge is signalled through the grammatical keywords "they" and "them", which Titus uses to objectify his enemies. After promising to "o'er reach *them* in their own devices" (5.2), he instructs a mutilated Lavinia to get "*them* [Chiron and Demetrius] ready" for the banquet (5.2). The latter equates to a grotesque perversion of a Eucharistic feast, with Tamora's sons served up to her in a pie (hence the keyword, "eat"). Titus appears to believe he was, albeit violently, righting wrongs as opposed to acting villainously: "*They* ravished [Lavinia], and cut away her tongue, / And *they*, t'was *they* that did her all this wrong", (5.3).

This inability to see himself as wrong-doer or to show remorse (cf. Charney 2012) is what ultimately problematizes Titus's hero status, in our view.

7. Discussion of the *King Lear* characters

The villainy in *King Lear* comes about because of the family divisions Lear triggers when dividing his kingdom. Lear wants his daughters – Goneril, Regan and Cordelia – to make public declarations of their love in return for portions of his kingdom. Goneril and Regan comply and receive land for themselves and their husbands. Cordelia's refusal to engage in the same way sees her disowned. Gloucester's bastard son, Edmund, meanwhile, deceives his father into believing Edgar (his older legitimate brother) is trying to kill Gloucester for his land. Edmund is named heir in consequence. Edmund then betrays his father to Regan (when Gloucester offers help to, first, Lear and, then, Cordelia) and pursues romantic relationships with Regan as well as Goneril (with the intention of cementing his power further).

7.1 Lear

When studied in their context-of-use, Lear's keywords confirm that the villain/hero dichotomy is far too simplistic for a protagonist who claims he is "a man / More sinned against than sinning" (3.2). Lear's statistical overuse of the keywords "Daughter", "she" and "her" is unsurprising given the plot. Lear draws upon "she/her" to emphasize Cordelia's (decreased) transactional worth after she fails to flatter him as her sisters had done: "When *she* was dear to us, we did hold *her* so, / But now *her* price is fallen" (1.1). It is here, too, Lear draws on his final use of the keyword "dower", to warn Cordelia's suitor, Burgundy, that she is now 'Unfriended, new adopted to our hate, / *Dowered* with our curse," before informing him to "Take *her* or, leave *her*". The above in conjunction with Lear's disowning of Cordelia – by calling her his "sometime *Daughter"* (1.1) – are so negatively

loaded/derogatory that they problematize Ray's (2007: 98) assessment of Lear as the "undisputed hero of the play". Lear's uses of the keyword "cause" can be understood as self-centered markers for his egotistical nature: a characteristic he displays for most of the play. It is only when he is reunited with Cordelia that he begins to (be able to) see from the perspective of others. He appreciates that Cordelia has "some *cause*" (4.6) for wanting to do him wrong, for example, while his other daughters "have not." Lear's uses of "she/her" in his attempt to revive Cordelia's dead body are further indications of his remorse at being ultimately responsible for her demise. He desperately hopes "*she* lives" while acknowledging "*she's* dead as earth" and "I might have saved *her*, now *she's* gone for ever" (5.3).

Lear's use of "she/her" in reference to Goneril is always negative, often to an extreme: "If *she* must teem, / Create *her* child of Spleen" (1.4). When Goneril reduces Lear's retinue of knights, he regards this particular "*Daughter*" as "a disease" which contaminates his "flesh" (2.2). He admits, in addition, to being "ashamed" Goneril had the "power to shake [his] manhood", causing him to shed "hot tears" that, from his perspective, she did not deserve. He then draws on the keyword "ha", but as a surge feature designed to heighten his rejection of her: "*Ha*? Let it be so. I have another daughter" (1.4). "Regan" is the only named daughter to be a keyword for Lear. Most mentions occur in Act 2, Scene 2 where Lear is (still) hopeful she will be more sympathetic towards him than Goneril: "Beloved *Regan* , / Thy Sister 's naught: oh *Regan*...., thou wilt not believe ... O *Regan*" (2.2).

The eight instances of the keyword "weep" again reveal something about Lear's transformation (or development) as a character. Six instances are accompanied by negation and/or "I", and demonstrate Lear's egotistical self-preoccupation for much of the play. He

states “I'll not weep”, in spite of having “full cause of weeping” for example (2.2, see also 3.4, 4.5 and 4.6). Once reunited with Cordelia, though, he focuses on the grief he has caused her: “I pray weep not” (4.6) and determines that when they are imprisoned together, their captors will starve “ere they shall make us weep” (5.1). The keyword is replaced by Lear's howls and tears when he enters with the corpse of Cordelia who has been hung in the prison, thus marking his maturation to tragic hero.

7.2 Edmund

Edmund's keywords in context modify Quennell and Johnson's view of Edmund's “self-awareness and delight in his own villainy” (2013: 63). Four of Edmund’s keywords – “base”, “legitimate”, “brother” and “father” – point to his obsession with social status and, crucially, his desire to achieve legitimacy or get “to the *Legitimate*” (1.2) in order to escape from the role of “base” villain.. He muses over why he is considered “base”, unworthy and inferior such that he cannot inherit his father’s estate – “Why Bastard? Wherefore *base*?” (1.2); “With baseness Bastardy? *Base*, *Base*?” (1.2) – before going on to plot against his brother, Edgar, stating – “Well then, / *Legitimate* Edgar, I must have your land…fine word: *Legitimate*. / Well, my *Legitimate*, if this Letter speed, / And my invention thrive, Edmund the *base* / Shall to the *Legitimate”* (1.2). As the previous turn reveals, Edmund’s plan to “have lands *by* wit” “if not *by* birth” (1.2) depends on projecting his own villainy (and social inferiority) onto “a *Brother* Noble” (1.2) who does not suspect Edmund's own “villainous” plotting that transforms Edgar into a base outcast. Edmund's success depends, in turn, upon “A Credulous *Father*,” Gloucester, who is ready to believe Edmund's tale that Edgar is the “villain” (1.2) trying to seize his father’s estate. Edmund subsequently plots to have Gloucester apprehended as a “villain” to the state (3.7) in order to advance his own rise to legitimate power.

"Sword" is also a keyword for Edmund, but this is because he uses it as a key prop in the play. He persuades his father of Edgar's supposed treachery by cutting himself with a sword (1.2); he attempts to defend his usurped position as Gloucester's noble, legitimate heir by duelling with the "villain-like" Edgar (5.3), and; prior to dying, attempts (in vain) to reprieve Lear and Cordelia (5.3) using his sword as a symbol. This last endeavour and his pathetic claim that "Edmund was beloved" (5.3) are marks of nobility / legitimacy that are his real goal and save him from being a one-dimensional villain that can only gloat about his evil "practices" over his brother and father (cf. Charney 2012).

8. Discussion of the *Macbeth* characters

Villainy in *Macbeth* is triggered by prophecies given by the three "weird sisters" who tell Macbeth he will become Thane of Cawdor, and then, King of Scotland. Macbeth conspires with Lady Macbeth to murder King Duncan, in consequence, but is unable to prevent Banquo's sons being the country's future kings (as the sisters foretold).

8.1 Macbeth

Macbeth's tentative claim to the throne throughout the play is indicated by his grammatical keyword "if", first seen when he contemplates assassinating Duncan:

> *If* it were done when 'tis done, then 'twere well
> It were done quickly: *if* the assassination
> Could trammel up the consequence, and catch
> With his surcease success […]
> We'd jump the life to come (1.7).

When the conditional is drawn upon again, it is as part of the fateful question – *"If* we should fail?" (1.7) – and, hence, serves to indicate Macbeth's emerging insecurity about sovereignty. This, in conjunction with the weightiness of Macbeth's growing guilt, accounts for the keyword

"fear". Macbeth confesses to "*fear*[ing]…Banquo", in particular. When Banquo's son Fleance escapes Macbeth's murder plot, his "doubts and *fears*" (3.4) intensify further. The keyword, "blood", a verbal equivalent to Lady Macbeth's incessant washing of her hands, signifies Macbeth's "agoniz[ing] with himself over his ill-doing" (Charney, 2012: 86). He acknowledges "I am in *blood* / Stepped in so far that, should I wade no more / Returning were as tedious as go o'er" (3.4), for example. Such guilt gives Macbeth "a special status in Shakespeare as a villain-hero", according to Charney (ibid). His haunting by an interminable line of Banquo's descendants means he is doomed to continue his murderous course. The keyword, "&/and", is repeated when Macbeth recalls a "seventh", an "eighth" and "many more" kings in a line stretching out to the "crack of doom", while "Banquo smiles upon [him] / *And* points to them" (4.1). It echoes alongside the keyword, "tomorrow" in Macbeth's nihilistic vision of the future following his wife's suicide: "Tomorrow, *and* tomorrow, *and* tomorrow… *And* all our yesterdays have lighted fools / The way to dusty death" (5.5). The six instances of the keyword "born" are all used as part of a phrase denoting a caesarean section (5.3, 5.7, 5.8), and (as Macbeth learns), refer ultimately to Macduff, the thane who slays him at the play's end, thus fulfilling the final part of the weird sisters' prophecy.

8.2 Lady Macbeth

Charney (2012: 86) is confident that Lady Macbeth is not a villain, in spite of having "murderous and savage thoughts". The keywords suggest a slightly different interpretation: that of a villain who is then haunted by the error of her ways. The keywords, "would" and "without", capture her manipulative disposition, for example. "Would" collocates with "thou" (x6) and "yet" (x3), the latter of which is also used with "without" when Lady Macbeth considers the likelihood of her husband taking the necessary steps to make himself king. She thinks he "*would'st* be great" and is "not *without* ambition", and "*wouldst* [aim] highly", for

example, but also feels he "*wouldst* not play false" and "*wouldst*" "win" "holily" because he is "too full o'th milk of human kindness" (1.5). Lady Macbeth's symbolically feminine keyword "without" expresses her own overwhelming sense of lack. She longs for the spirits to "unsex me here", and lambasts Macbeth for any sign of feminine weakness: "My hands are of *your* colour; but I shame / To wear a heart so white" (2.2). In the banquet scene, she scolds him for being "unmanned by folly" (3.4) when he sees the ghost of Banquo. Like her husband, though, Lady Macbeth ultimately realizes they are trapped by lack: "Naught's had; all's spent / When our desire is got *without* content", and can only live in negative terms: "things without all remedy / Should be without regard" (3.2) Her sleepwalking scene (5.1), rhythmically repeats the keyword "bed" amongst other negatives: "To bed, to bed ... What's done cannot be undone. To bed, to bed, to bed" (5.1.) It inverts the earlier murder scene where she urges Macbeth to wash his hands and get to bed, even though he can "sleep no more". (2.2). Ironically, it is now the sleeping Lady Macbeth who is without - as in outside - the bed and cannot rest. With hindsight, she inverts the ambitious impulse that guided her earlier use of the conditional "would" to lament "who / Would have thought the old man to have had so much / Blood in him". This unintended confession ultimately leads to a public condemnation of the Macbeths as villains: "this dead butcher and his fiend-like queen". Yet, such a judgement seems hugely reductive, based on the keywords in context.

9. The seven Shakespearean characters: hero, anti-hero or villain?

Our aim in this paper was to demonstrate how a corpus linguistic technique like keyword analysis can benefit – by confirming/refuting or advancing – existing literary understandings of Shakespearean depictions of behaviour deemed villainous. Our keyword results were gleaned using a cutting-edge, three-step filtering process that leads to fewer but arguably more robust results, quantitatively and qualitatively speaking, than would be achieved by relying on

LL alone. Each of the generated keywords was then checked (via concordance lines) to reveal those that, when studied in context, tell us something about the characters' malicious / heroic qualities. Aaron proves to be the only character of the seven to have "villainy" as a keyword, though Edmund cleverly projects his own villainy onto others, driven by the patrilineal system's exclusion of him as base born (the original meaning of villain). Our analysis of keywords-in-context shows that both these base-born villains are more complex than the Vice figures of medieval morality plays. Aaron's use of "villainy" along with the keywords "black" and "gold" reveal his ambition to challenge the stereotypical connections between race and villainy (in the sense of both social inferiority and evil behaviour) for his son, if not for himself. Edmund's keywords "base" and "legitimate" reveal he is a villain driven by a desire to attack, albeit with the ultimate intention of inserting himself in "to the legitimate," (1.2), to win a place in the patrilineal system which excludes him based on his birth.

Our keyword analyses likewise uncover much about the tragic heroines Tamora and Lady Macbeth. Tamora begins as a tragic victim but grows into a towering figure of revenge, modelled on classical feminine anti-heroes like Medea (albeit aided and abetted by Aaron's villainous plots). We have focussed on Tamora's use of "sweet", in order to highlight her use of speech to flatter and manipulate. Such a depiction makes her an early example of a characteristically feminine villainy (Tassi 2011; Pollard 2017). Lady Macbeth is as manipulative as Tamora, but her chosen methods involve "unsex[ing] herself" (Charney, 2012: 86) and emasculating Macbeth with the aid of keywords such as "would" and "without". As the play progresses, the "would" and "without" keywords also provide us with a window into this character's insecurity: in particular, the sense of lack that leads ultimately to her undoing herself through suicide and confession.

Titus is a military hero-turned-villain by tragic circumstances (that he inadvertently set in motion himself). We see him as a villain, more than hero (or even villain-hero), in spite of his ultimately shedding "tears", due to his inability to see himself as wrong-doer/show remorse (cf. Charney 2012). King Lear also sheds tears. Unlike Titus, however, his early cruelty to Cordelia and egotistical indifference to others is slowly transformed as signalled by his keywords "she"/"her"; "cause" "weep" and its collocates, "I" and "not". The suffering he endures at the hands of others, along with his increasing acknowledgement of his folly and responsibility make Lear "a man more sinned against than sinning" (3.2), meaning that even the villain-hero label is too simplistic to account for his growth in tragic awareness. Macbeth is arguably a victim as well as agent of villainy. Charney (2012: xvii) describes him as a "villain-hero" because he is perpetually tormented by his own guilt; and this overwhelming sense of anxiety/dread (if not guilt) is evidenced by the keywords "if" and "fear".

When taken collectively, we believe that such results provide a convincing argument that keyword analyses can illuminate the linguistic patterns that give nuance to characters whose actions are morally reprehensible or questionable, at the same time as grounding previous (literary) understandings of these characters' thoughts, feelings and intentions in (empirical) linguistic analysis.

References

Archer, D. and D. Bousfield. 2010. "See better, Lear"? See Lear better! A corpus-based pragma-stylistic investigation of Shakespeare's *King Lear.* In *Language and Style*. Daniel McIntyre and Beatrice Busse (eds), 83–203. London: Palgrave.

Archer D. and C. Lansley. 2015. Public appeals, news interviews and crocodile tears: an argument for multi-channel analysis. *Corpora* 10(2): 231–258.

Archer, D., J. Culpeper and P. Rayson. 2009. Love – 'a familiar of a devil'? An Exploration of Key Domains in Shakespeare's Comedies and Tragedies. In *What's in a word-list? Investigating word frequency and keyword extraction*. D. Archer (ed) 137–158, Farnham, England: Ashgate.

Baron, A. and P. Rayson. 2008. VARD2: A tool for dealing with spelling variation in historical corpora. *Postgraduate Conference in Corpus Linguistics*, Aston University, 22 May 2008.

Bowers, F.T. 2015 [first ed. 1959]. *Elizabethan Revenge Tragedy, 1587-1642.* New Jersey, USA: Princeton University Press.

Busse, B. 2006. *Vocative Constructions in the Language of Shakespeare*. Amsterdam and Philadelphia: John Benjamins.

Busse, U. 2002. *Linguistic Variation in the Shakespeare Corpus*. Amsterdam and Philadelphia: John Benjamins.

Charney, M. 2012. *Shakespeare's Villains.* Madison: Fairleigh Dickinson University Press.

Culpeper, J. Forthcoming. General Introduction. *Encyclopedia of Shakespeare's Language*. [Arden Shakespeare] London: Bloomsbury.

Culpeper, J. 2009. Keyness: words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics* 14(1): 29–59.

Culpeper, J. 2002. Computers, language and characterisation: An Analysis of six characters in Romeo and Juliet. In *Conversation in Life and in Literature: Papers from the ASLA Symposium* [Association Suedoise de Linguistique Appliquee 15], U. Melander-Marttala, C. Ostman and M. Kytö (eds), 11–30. Uppsala: Universitetstryckeriet.

Garside, R. and N. Smith. 1997. A hybrid grammatical tagger: CLAWS4. In *Corpus Annotation: Linguistic Information from Computer Text Corpora,* R. Garside, G. Leech and A. McEnery (eds), 102–121. London: Longman.

Oxford English Dictionary, 2nd. rev. ed. (nd). Oxford. Oxford University Press.

Pollard, T. 2017. *Greek Tragic Women on Shakespearean Stages.* Oxford: Oxford University Press.

Quennell, P. and H. Johnson. 2013. *Who's Who in Shakespeare*. London and New York: Routledge.

Ray, R. 2007. *William Shakespeare's King Lear.* [Atlantic Critical Studies] New Delhi: Atlantic Publishers.

Rayson, P., D. Archer, A. Baron, J. Culpeper and N. Smith. 2007. Tagging the Bard: Evaluating the accuracy of a modern POS tagger on Early Modern English corpora. In *Proceedings of the Corpus Linguistics Conference: CL2007*, M. Davies, P. Rayson, S. Hunston and P. Danielsson (eds) University of Birmingham, UK, 27–30 July 2007.

Tassi, M.A. 2011. *Women and Revenge in Shakespeare: Gender, Genre and Ethics.* Selinsgrove: Susquehanna University Press.

White, J.S. 1997. "Is black so base a hue?": Shakespeare's Aaron and the Politics and Poetics of Race. *College Language Association Journal* 40: 336–66.